The Quarterly Journal of Experimental Psychology

ISSN: 0272-4987 (Print) 1464-0740 (Online) Journal homepage: http://www.tandfonline.com/loi/pqja20

Grouping frequency components of vowels: When is a harmonic not a harmonic?

C. J. Darwin & N. S. Sutherland

The Quarterly Journal of Experimental Psychology (1984) 36A, 193–208

Grouping Frequency Components of Vowels: When Is a Harmonic Not a Harmonic?

C. J. Darwin and N. S. Sutherland
Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, U.K.

When one harmonic of a vowel starts before and stops after the others, its contribution to the vowel's phonetic quality is reduced. Two experiments demonstrate that this reduction cannot be attributed entirely to adaptation. The first experiment shows that a harmonic that starts at the same time as a short vowel but continues after the vowel has ended contributes almost as little to the vowel's phonetic quality as a harmonic that starts before but stops at the same time as the vowel. The second experiment shows that the small contribution to vowel quality of a harmonic that starts before a vowel can be increased by adding an additional tone that will in turn form a perceptual group with that part of the harmonic preceding the vowel. The experiments demonstrate that some perceptual grouping operations are performed before the first formant of a vowel is estimated from the amplitudes of its component harmonics.

GENERAL INTRODUCTION

We normally listen to speech against a background of other sounds. Yet almost all experiments on speech perception investigate how we perceive speech when it is the only sound present. Such experiments have given us considerable knowledge about how linguistic categories are perceived from the speech of a single speaker, but they have led to perceptual models whose performance would not degrade gracefully in the presence of other sounds. For example, Klatt's (1980a) laudably explicit model of acoustic–phonetic analysis and lexical access matches the incoming sound against a set of static spectral templates. The model implicitly assumes that all the input sound is relevant. Consequently, the template that would be

Requests for reprints should be sent to C. J. Darwin, Laboratory of Experimental Psychology, University of Sussex, Brighton, Sussex, BN1 9QG, U.K.

S. D. Isard and H. C. Longuet-Higgins helped us to think about the experiments. John Doyle, who ran the experiments, and John Seton, who wrote some of the computer programs, were supported by a grant to Sutherland from the SERC, which also provided the computer facilities.

the best match to an isolated speech sound would not necessarily be the best when other sounds are present. It is clear that human speech perception is not so delicate. Our ability to separate sounds that emanate from different sources was studied in the 1950s, predominantly in the context of selective attention (Miller and Heise, 1950; Cherry, 1953; Broadbent, 1958), and now continues against a background of work on auditory perceptual grouping (Bregman, 1978).

Bregman and his colleagues have investigated the principles by which the auditory system organises or parses its input into "streams", corresponding to separate putative sources of sound. Just as in vision there are grouping principles that determine, for example, figure and ground, so in hearing there are principles that guide the auditory system in deciding which of the frequency components present at a particular moment have originated from a common sound source (for reviews see Bregman, 1978; McAdams and Bregman, 1979).

Experiments of our own (Darwin, 1981) have examined the problem of which of the formants present at a particular time are grouped together for the purpose of phonetic categorisation. They extended earlier work (Broadbent and Ladefoged, 1957) in demonstrating that formants whose harmonics share a common fundamental were more likely to be grouped together into a phonetic percept than formants whose harmonics were on different fundamentals. They also showed that formants that started at different times were, under some circumstances, less likely to be grouped together than formants that started at the same time.

The experiments reported here look at what may be an earlier stage in speech processing, the estimation of formant frequencies themselves. A formant frequency is a resonant frequency of the vocal tract, appearing in the spectrum of a speech sound as a peak in the spectral envelope. Figure 1 shows spectra of the first formant region of three vowels that differ in their first formant (F1). They all have the same fundamental frequency (125 Hz), so all the frequency components are integer multiples of 125 Hz. The vertical lines represent the frequency components, while the dotted line shows the spectral envelope. The peak in the envelope lies at the formant frequency. At the top left of Figure 1 the spectrum is of a vowel synthesised with a first formant of 375 Hz. Here the first formant frequency coincides with the third harmonic. At top right is the spectrum of a vowel with a first formant at 500 Hz, coinciding with the fourth harmonic. Between them lies the spectrum of a vowel with a first formant at 460 Hz. Now the formant peak does not coincide with a harmonic, but lies between the third and the fourth. It is clear that a change in formant frequency leads to a change in the relative levels of harmonics that lie close to the formant in frequency. The first formant frequency is an important determinant of the perceived phonetic quality or colour of a vowel; but how is the first formant estimated from the intensities of the individual harmonics? It is clear that the system does not simply take the

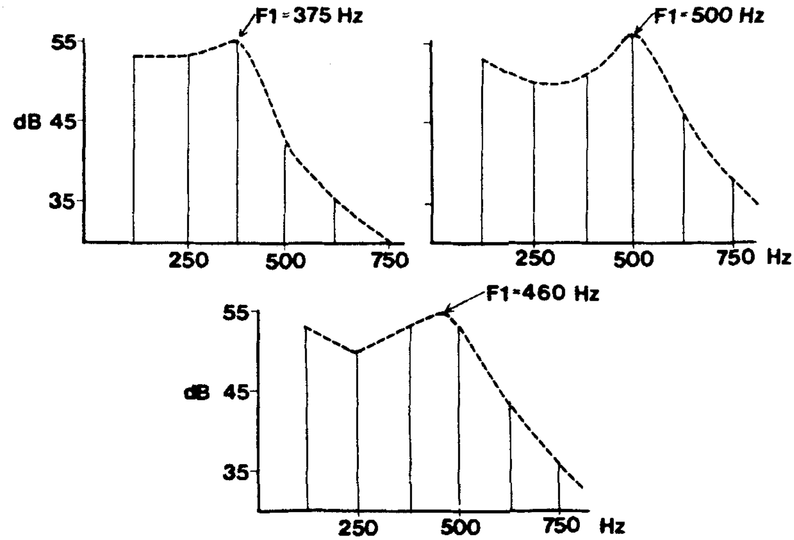

Figure 1. Spectra of the first formant region of three vowels differing in the frequency of their first formant. The two upper spectra are taken from the end-point vowels of the basic continuum between /I/ and /e/ used in the two experiments. The lower spectrum illustrates a vowel close to the phoneme boundary.

formant frequency to be the frequency of the most intense harmonic; rather, it appears to estimate the first formant frequency by, for example, forming a weighted sum of the harmonic frequencies near the formant, with the weights corresponding to the harmonics' intensities (Karnickaya, Mushnikov, Stepokurova and Zhukov, 1975; Carlson, Fant and Granstrom, 1975).

The spectra in Fig. 1 are for isolated vowel sounds. But if the vowel had been produced against a background of other sounds, the spectrum would have contained components from other sound sources. How is vowel quality influenced by the presence of other sounds? What information does the perceptual system use to separate those frequency components that come from the vowel from those that are extraneous? The system may group together frequency components that are harmonically related (Broadbent and Ladefoged, 1957; Brokx and Nooteboom, 1982; Darwin, 1981) or which change in frequency together (c.f. McAdams and Bregman, 1979; McAdams, 1980), or which start and stop at different times (cf. Bregman and Pinker, 1978; Dannenbring and Bregman, 1978).

The experiments described below follow up a recent finding (Darwin, 1983) that a harmonic that starts and stops at a different time from the rest of a vowel makes a reduced contribution to the vowel colour. A difference in onset and offset time of only 20 or 30 msec is sufficient to reduce the contribution of a harmonic to the overall vowel colour. It is also sufficient to make the tone audible as a separate tone (cf. Rasch, 1978).

In this paper we distinguish two types of explanation for the reduced contribution of a leading harmonic to vowel colour: adaptation and grouping. According to an adaptation explanation, processes sensitive to energy at the frequency of the leading harmonic adapt during the time it is present before the main vowel starts, so that at the onset of the main vowel energy at the tone's frequency makes a reduced contribution to the overall timbre. Adaptation with a sufficiently rapid time course does in fact occur in primary auditory neurone responses to tone bursts and speech sounds (e.g. Delgutte, 1980; 1982) and is probably due to depletion of neuro-transmitter in hair cells (Schwid and Geisler, 1982). Adaptation is a possible explanation for at least some aspects of a variety of auditory temporal phenomena (Zwicker, 1964; Cardozo, 1967; Neelen, 1967; Wilson, 1970; Lummis and Guttman, 1972; Viemeister, 1980; Viemeister and Bacon, 1982; Summerfield, Foster, Gray and Haggard, 1981) and is an attractively simple solution to the problem of grouping together sounds that start at the same time.

By contrast, a grouping explanation (cf. Bregman, 1978; Bregman and Pinker, 1978; Dannenbring and Bregman, 1978) allows perceptual mechanisms to separate the internal representation of the leading tone from the other frequency components, by virtue of its different onset time. An important difference between these explanations is that while a tone that has been removed from a complex by perceptual grouping could be reinstated by alternative groupings, one that has adapted could not.

Darwin (1983) attempted to distinguish between these two explanations by comparing the effectiveness of onset and offset time differences. If a harmonic that starts at the same time as vowel but stops at a different time makes a reduced contribution to the vowel colour, it cannot do so by adaptation. Some marginal evidence was obtained for an effect of offset time, but the size of the effect was very small compared with that of onset-time differences

A plausible reason for the minimal success of that experiment is that it used relatively long vowels (320 msec). The subject could well have decided on the phonetic category of the vowel before hearing the lagging tone that followed it. The first experiment to be reported examines the effect of leading and lagging tones on the phonetic quality of short (50-msec) vowels and does indeed find a sizeable effect of tone offset. The second experiment uses a different technique to reinforce the conclusion that adaptation is not an adequate explanation for the changes in vowel colour produced by a leading tone.

EXPERIMENT 1

We have already shown that a harmonic that starts and stops at different times from the rest of a vowel makes less contribution to the vowel's quality than one that starts and stops at the same time. The effect could be due to adaptation. If it is *entirely* due to adaptation, then a tone that starts at the same time as the vowel but continues after the vowel has stopped should contribute just as much to the vowel as a tone that starts and stops at the same time. On the other hand, the effect could be due to perceptual grouping. Differences in onset and offset time could be used to group frequency components differentially. Any effect of offset time could be explained by grouping, but not by adaptation.

The previous experiment (Darwin, 1983) found only marginal evidence for an effect of offsets, but it used rather long vowels. The present experiment uses short vowels to give offsets a better chance of exerting an influence.

Method

The technique used to estimate the phonetic quality of different sets of vowels was to find the boundary between the vowels /I/ and /e/ along a continuum that differs in first formant frequency. The boundary for the basic continuum is estimated and compared with boundaries obtained from a variety of other continua that have been derived from the basic one by adding extra energy to one of the harmonics. Adding extra energy changes the vowel quality along the /I/–/e/ continuum. Let us refer to each sound in the new continua as having the same nominal first formant frequency as its parent sound in the original continuum. We can now measure the change in vowel quality as a shift in the nominal first formant frequency at the phoneme boundary. If the extra energy is made to start earlier than the main vowel, the phoneme boundary moves back towards the original position. The position of the phoneme boundary relative to its original value thus gives a measure of the contribution that the extra energy is making to the vowel quality.

Stimuli

The basic continuum had seven sounds, whose first formants differed in equal (Hz) steps between 375 and 500 Hz. The vowels were synthesised on a constant fundamental frequency of 125 Hz, so that the first formant fell between the third and fourth harmonics. The first formant bandwidth was fixed at 70 Hz. The second through fifth formants were fixed at 2300, 2900, 3800 and 4600 Hz. The vowels were 56 msec long, including a 16-msec rise and fall time, so that the steady-state was 24 msec (or 3 pitch pulses).

Nine more continua (each with seven members) were constructed from the basic continuum by adding various 500-Hz tones to each member. Each 500-Hz tone had twice the amplitude of the 500-Hz component of the vowel to which it was added. Since this tone had the same phase as the original 500-Hz component, the new vowel had 9.5 dB more energy at 500 Hz than the original vowel. Schematic waveform envelopes and spectra for a member of the original vowel continuum, and the corresponding sound with extra energy starting and stopping 32-msec from the vowel, are shown in Fig. 2.

The various continua differed in the duration and temporal alignment of the added tone. In the simplest case, the additional tone had the same duration as the vowel to which it was added and also started and stopped at the same time. In other cases the tone started and/or stopped 32 or 240 msec from the vowel. All nine possible combinations of three (0-, 32- and 240-msec) onset and offset times were used. The spectra of the resulting vowels were checked to ensure that the

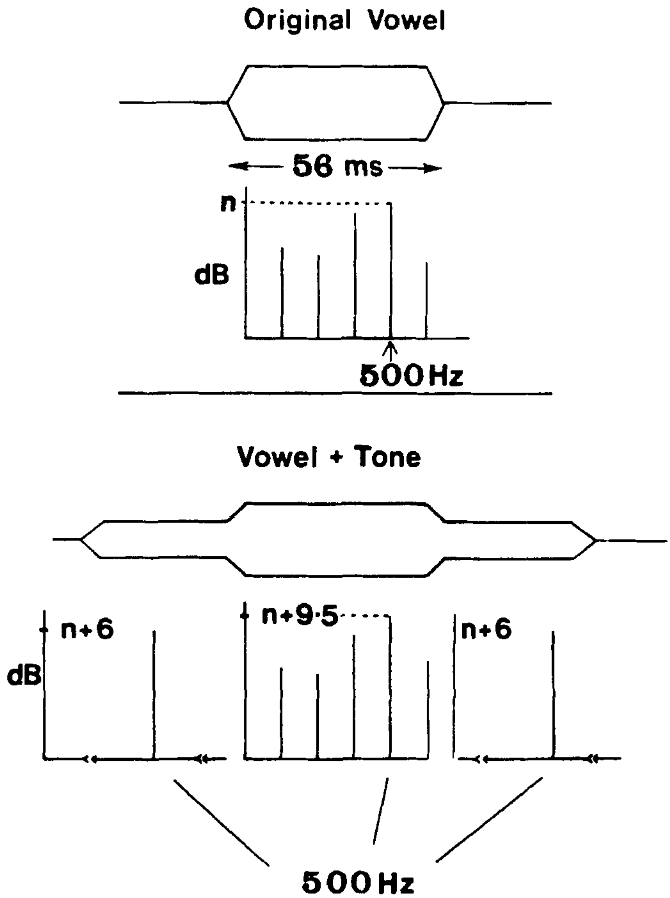

Figure 2. Schematic waveform envelopes and line spectra for two of the conditions in Experiment 1. The upper part of the figure shows one of the vowels from the original continuum. The lower part of the figure shows the corresponding vowel where extra energy has been added at 500 Hz. In the experiment, onset and offset times for the tone were set independently to 0, 32 or 240 msecs.

phase relations between the additional tone and the vowel stayed constant to produce the required increase in level.

Synthesis Techniques

The basic continuum of sounds was synthesised using a software serial-formant speech synthesis program (Klatt, 1980b). The additional 500-Hz tones were produced for each member of a continuum by filtering an appropriate vowel twice through a 101 coefficient FIR filter, which attenuated harmonics other than 500 Hz by at least 56 dB. The filtering introduced a time delay of 10 msec, so before adding in the filtered tone the original vowel was also shifted by 10 msec.

Procedure

Six subjects, who were research students and staff of the university, were tested individually in a sound-proof booth over Sennheiser HD-414 headphones. The level of the lowest first formant member of the basic continuum was approximately 58 dB(A).

The sounds were produced on-line by the laboratory's VAX 11/780 computer (via an LPA-11K at a sampling frequency of 10 kHz, low-pass filtered at 4.5 kHz and 48 dB/oct). Subjects responded on a conventional VDU keyboard with either of the two identification responses ("I" for /I/ and "E" for /e/). If they were not sure of a sound's category on first hearing, they could press "R" to repeat it. Following each key-press, the appropriate next sound was played after a pause of 1 sec. The VDU screen gave information on allowed responses and the current trial number, as well as castigating the use of other keys.

Each subject heard 10 successive random sequences of the 70 items (10 conditions × 7 continuum steps) and was free to take a rest at any time. The main experiment was preceded by a demonstration of the basic continuum and the continuum with the synchronous tone added (but without telling the subject that a tone had been added), followed by a practice identification session using 10 successive random orderings of those two continua.

After the practice session subjects were told that they might now hear tones mixed in with the vowels, but they were to ignore them and simply report the vowel that they heard.

Results

The pooled identification functions for the six subjects and for six of the nine continua are shown in Figures 3 and 4. They are expressed as the

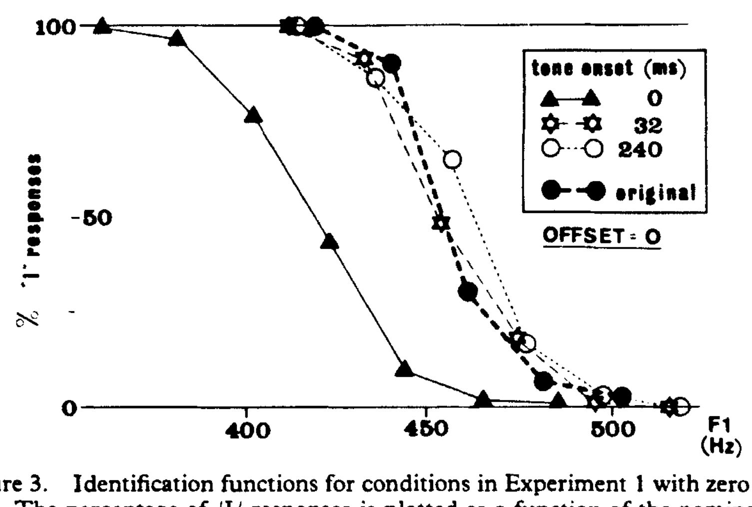

Figure 3. Identification functions for conditions in Experiment 1 with zero offset time. The percentage of /I/ responses is plotted as a function of the nominal first formant frequency. Different conditions have extra energy at 500 Hz starting at different times but stopping at the same time as the vowel. The nominal first formant frequency is the frequency used to synthesise the sound from the original continuum, before extra energy was added.

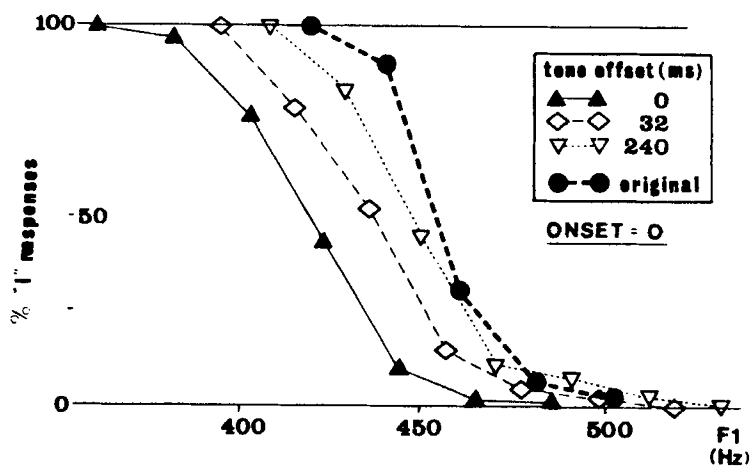

Figure 4. As Figure 3, but plotting conditions in Experiment 1 with zero onset time.

percentage of /I/ responses given to each sound, where each sound is referred to by the first formant frequency of the sound from the basic continuum from which it was derived. If additional energy made no difference to vowel quality, the phoneme boundary would be at the same nominal first formant frequency. In order to pool across subjects without confounding between-subject variability in boundary position with the slope of the individual identification functions, the curves of the individual subjects were aligned around a common boundary before averaging. The resulting curve was then plotted at the average boundary. The slope of each plotted curve thus gives the average slope of the individual identification functions.

The phoneme boundaries for each subject in each condition were found by probit analysis, and the average phoneme boundaries are shown in Figures 5 and 6 for all ten conditions. In Figure 5 they are plotted as a function of onset time, with offset time as the parameter. In Figure 6 onset time is the parameter.

The main findings are as follows. First, when the additional energy starts and stops at the same time as the vowel, there is a clear shift in vowel quality, so that the phoneme boundary moves to a lower nominal F1 frequency. Second, when the additional tone starts before the vowel or stops after it, or does both, the original vowel quality is heard. A 32-msec onset shifts the boundary significantly from the simultaneous condition [$t(5) = 6.1$, $p < 0.001$] so that it is not significantly different from that of the basic continuum. A 32-msec offset also shifts the boundary significantly away from the simultaneous condition [$t(5) = 3.03$, $p < 0.05$], although it is still different from the basic continuum's boundary [$t(5) = 2.97$, $p < 0.05$]. With a 240-msec offset, the original vowel colour

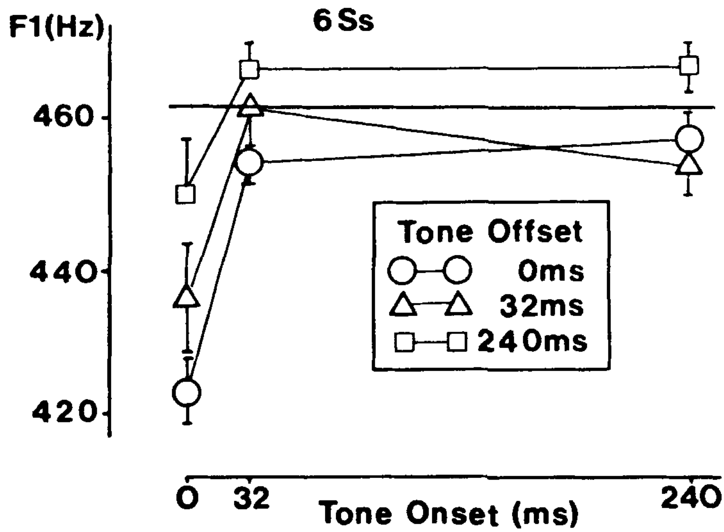

Figure 5. Nominal first formant values at the /I/–/ɛ/ phoneme boundary for vowel sounds in Experiment 1. The horizontal line across the figure is the boundary for the basic continuum. The remaining points denote the boundary for sounds that have had extra energy added to their fourth harmonic (500 Hz). The formant values are those used to synthesise the original continuum. The ordinate gives the number of msec by which the extra tone precedes the vowel. The parameter is the number of msec that the extra tone continues after the vowel. The values are means (and standard errors) across 6 subjects.

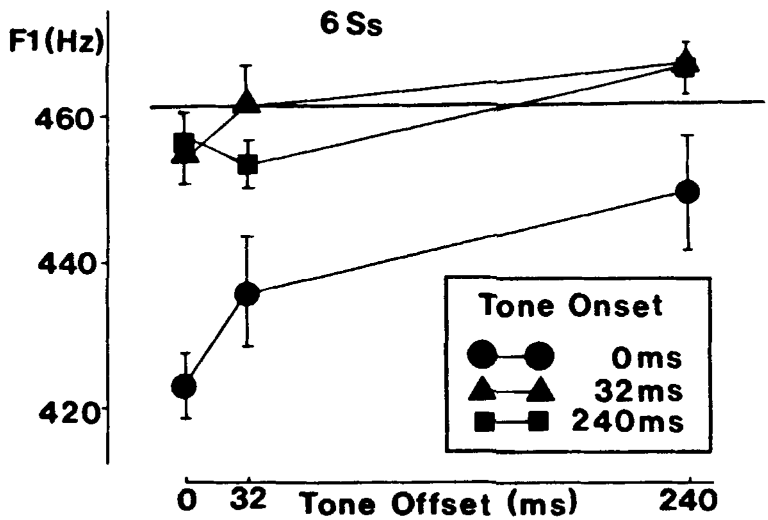

Figure 6. The same data as in Figure 3, plotted with the ordinate and parameter dimensions reversed.

is heard [$t(5) = 1.34$, $p < 0.1$] with a highly significant shift in boundary from the simultaneous condition [$t(5) = 5.79$, $p < 0.001$]. The difference between the 32-msec onset condition and the 32-msec offset just falls short of significance [$t(5) = 2.50$, $p < 0.055$], but it is likely that this difference could be established by increasing the number of subjects.

Discussion

The experiment replicates our previous finding (Darwin, 1983) that a tone added to a vowel at one of its harmonic frequencies will contribute less to the vowel's quality when it *starts* before than when it is simultaneous with the vowel. This effect could be due either to adaptation or to perceptual grouping. The experiment also demonstrates an effect that cannot be due to adaptation, namely, that a tone that *stops* at a different time from a vowel contributes less to its quality than does one that is simultaneous. Although this effect cannot be explained by adaptation, it can be explained by perceptual grouping.

Although it is rather implausible, it could be the case that the effect of onset is due entirely to adaptation, whereas the effect of offset is due to perceptual grouping. The next experiment demonstrates that a substantial part of the onset effect *is* due to perceptual grouping but leaves open the possibility that adaptation could contribute to the onset effect.

EXPERIMENT 2

The logic of the next experiment is as follows. A 500-Hz tone that starts before a vowel makes a reduced contribution to its quality. This may be because the leading part of the tone forms, with its continuation into the vowel, a separate perceptual group. If so, it should be possible to weaken the grouping between the leading part of the tone and its continuation by introducing a new tone (say at 1 kHz, a harmonic frequency of the leading tone) that starts at the same time as the leading tone, but stops *as the vowel starts*. The new tone might then form a perceptual group with the leading part of the 500-Hz tone, preventing it from grouping with its continuation into the vowel. The continuation would then form part of the vowel, resulting in a vowel quality closer to that heard when the 500-Hz tone is only present simultaneously with the vowel. A grouping explanation thus predicts that the contribution to vowel quality from a 500-Hz tone that starts before a vowel should be increased by introducing another tone that is synchronous with the leading portion of the 500-Hz tone.

Peripheral adaptation is frequency specific and would not be influenced by frequencies outside the critical band. Two-tone suppression effects

(Houtgast, 1974) of a higher tone on a lower have a similar scope and are negligible at the octave interval that we are using. Consequently, if the effect of a leading tone is due to adaptation, adding another tone synchronous with it but remote in frequency should not influence the vowel quality.

Method

Stimuli

Three of the continua from the first experiment and two new continua were used. Schematic narrow-band spectrograms for the five conditions are shown in Figure 7. The three continua from the previous experiment were (1) the basic continuum, (2) the continuum with extra energy at 500 Hz simultaneous with the vowel and (3) the continuum with extra energy at 500 Hz starting 240 msec before the vowel but stopping simultaneously with it. The two new continua were formed from (2) and (3) by adding to them an extra tone at 1 kHz at the same level as the 500-Hz leading tone, starting 240 msec before the vowel and stopping as the vowel started. To ensure that the 1-kHz tone had the same onset characteristics as the 500-Hz tone, it was produced in the same way, by filtering a vowel similar to those used previously (but with the second formant at 1 kHz in order to boost energy at that frequency) through a FIR filter at 1 kHz. The tone had 16-msec rise and fall times, and its fall occurred at the same time as the main vowel's rise, so that the extra tone had died away completely by the time the main vowel reached its steady-state level. The levels and timing of the various components of the sounds were checked by a digital spectrograph.

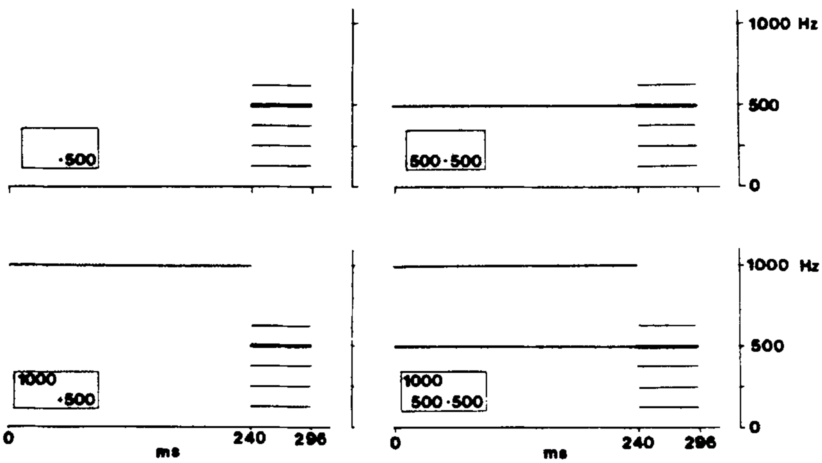

Figure 7. Schematic narrow-band spectrograms of single members of four of the stimulus conditions used in Experiment 2. In the top left-hand figure extra energy (denoted by a thicker line) has been added to the 500-Hz harmonic. In the top right-hand figure the additional energy starts before the vowel. In the two lower figures a 1-kHz tone has been added.

Procedure

Twelve paid subjects (students and staff of the university) who were not used to listening to synthetic speech were tested individually. The procedure was the same as in the first experiment, except that the total number of conditions in the main part of the experiment was reduced to 5, giving a total of 350 trials (5 conditions × 7 continuum steps × 10 tokens).

Results

Individual phoneme boundaries were calculated by probit analysis for each subject in each condition. The average boundaries across subjects are shown in Figure. 8. The three conditions brought across from the first experiment show similar results to that experiment. Adding energy at 500 Hz synchronously with the original vowel shifts the boundary [$t(11) = 11.2$, $p < 0.001$], but making the additional 500-Hz tone start 240 msec earlier significantly shifts the boundary back *towards* its original position [$t(11) = 10.0$, $p < 0.001$], although not quite *to* that position [$t(11) = 2.91$, $p < 0.01$].

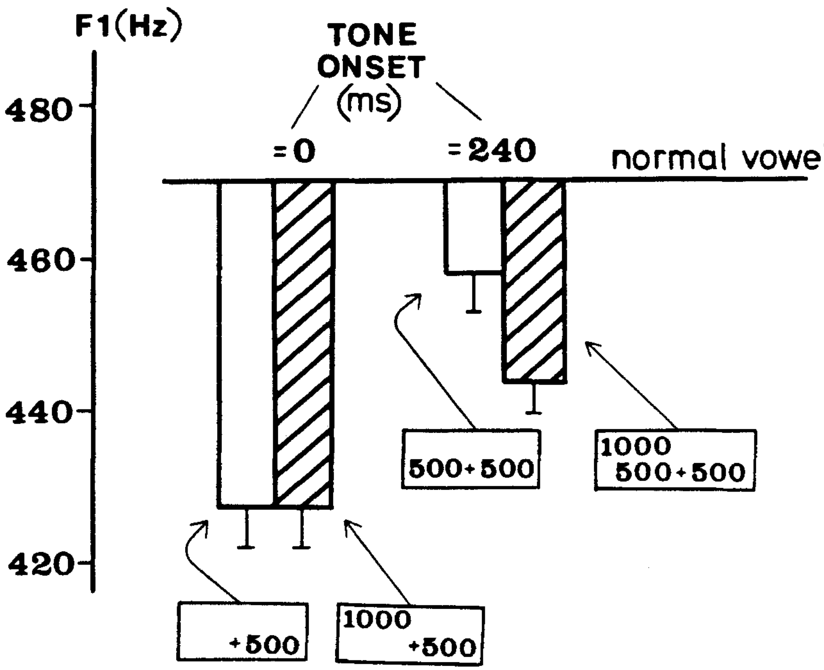

Figure 8. Mean (and their standard error) phoneme boundaries between /I/ and /e/ expressed as nominal first formant frequency for 9 subjects in Experiment 2. The boxed numbers refer to the conditions shown schematically in Figure 7.

The crucial test is whether adding the 1-kHz tone to the condition with the leading 500-Hz tone significantly shifted the boundary in the direction of the condition with no leading 500-HZ tone. It did [$t(11) = 5.54$, $p < 0.001$]. But adding it to the condition without the leading tone did not ($t = 0.4$), and analysis of variance showed this interaction to be significant [$F(1, 11) = 9.97$, $p < 0.01$]. However, adding the 1-kHz tone to the condition with the leading 500-Hz tone did not completely eliminate the effect of that tone, since the boundary was significantly different from its position with neither tone present [$t(11) = 7.00$, $p < 0.001$]. In fact, the boundary moved about a third of the way.

Discussion

The results of Experiment 2 clearly indicate that *not* all of the effect of a tone onset on vowel quality, found here and in Experiment 1, can be attributed to adaptation. At least half of the effect must be attributed to more complex processes such as perceptual grouping.

A 1-kHz tone that starts before a vowel and stops just as the vowel starts produces no change in the vowel's quality as measured by the phoneme boundary. But the 1-kHz tone *does* produce a change in vowel quality when it starts at the same time as a 500-Hz tone that continues into the vowel. The change in vowel quality is what would be expected if the two tones preceding the vowel formed a separate perceptual group that ends as the vowel starts, leaving the continuation of the 500-Hz tone into the vowel to be perceived as part of the vowel. When the 500-Hz tone precedes the vowel without the 1-kHz tone, although it continues into the vowel, it is perceived as separate from the vowel. If the corresponding change in vowel quality were due to adaptation around 500 Hz, it would not be counteracted by the presence of another tone at 1 kHz.

GENERAL DISCUSSION

We have demonstrated that the contribution made to vowel quality by a harmonic that starts at a different time from the others cannot be explained in terms of a simple peripheral process such as adaptation. The extent to which an individual harmonic contributes towards vowel colour depends not only on whether it starts at the same time as other harmonics, but also on whether it stops at the same time and on the extent to which it forms a perceptual group with other sounds. It is clear, then, that even such an early stage in speech perception as formant frequency estimation must (at least sometimes) be preceded by processes that assign frequency components to different apparent sources.

The dominant problem raised by our demonstrations is to determine the type of knowledge that is used by the grouping processes. In the experiments described here we have concentrated on the simple physical dimension of onset or offset time. It is a dimension that is undoubtedly of general use in grouping together sounds from the same source. But our demonstration introduces something of a paradox. If it is only harmonics starting and stopping at the same time that can be grouped perceptually to give a formant frequency, then virtually all natural speech would be unintelligible. The reason is that natural speech consists of formants that move rapidly in frequency; as a formant moves, it changes the amplitude of the harmonics around it. So a rising formant will progressively amplify higher harmonics and attenuate lower ones. The harmonics present at any one time will have become audible at different times and will become inaudible at different times, yet apparently we have no difficulty in hearing the timbre of the formant. Similarly, it is clear from pilot experiments we have made on vowels that have had a particular harmonic boosted by varying amounts that we do not need to incorporate all the harmonics that start at a particular time into a vowel's colour. Some increase from the "normal" level is interpreted as a change in formant frequency (as we have exploited in these experiments), but at some point additional energy is no longer incorporated into the vowel, and the immediate impression is then of a vowel with an extra tone present. Quite what determines the limits of this effect remains to be seen, but it may well reflect a property of speech sounds rather than a property of sounds in general.

Our experiments show that onset and offset time can be used to separate perceptually harmonics that do contribute to a vowel from those that do not. But such times constitute neither necessary nor sufficient conditions for grouping the harmonics of a single voice. Other principles must be used to ensure that the rapidly modulated harmonics of normal speech are grouped together, and that components occurring simultaneously by accident can be rejected.

REFERENCES

Bregman, A. S. (1978). The formation of auditory streams. In Requin, J. (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Lawrence Erlbaum Associates. Pp. 63–76.

Bregman, A. S. and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32, 19–31.

Broadbent, D. E. (1958). *Perception and communication*. Oxford: Pergamon.

Broadbent, D. E. and Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, 29, 708–710.

Brokx, J. P. L. and Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10, 23–26.

Cardozo, B. L. (1967). Ohm's law and masking. *IPO Annual Progress Report*, Vol. 2. Eindhoven, Netherlands: Institue for Perception Research. Pp. 59–64.

Carlson, R., Fant, G. and Granstrom, B. (1975). Two-formant models, pitch and vowel perception. In G. Fant and M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic. Pp. 55–82.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975–979.

Dannenbring, G. L. and Bregman, A. S. (1978). Streaming versus fusion of sinusoidal components of complex tones. *Perception and Psychophysics*, **24**, 369–376.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset time. *Quarterly Journal of Experimental Psychology*, **33A**, 185–207.

Darwin, C. J. (1983). Auditory processing and speech perception. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and performance X*. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory nerve fibers. *Journal of the Acoustical Society of America*, **68**, 843–857.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson and B. Granstrom (Eds.), *The representation of speech in the peripheral auditory system*. Amsterdam: Elsevier Biomedical Press.

Houtgast, T. (1974). *Lateral suppression in hearing*. Doctoral dissertation, Free University of Amsterdam.

Karnickaya, E. G., Mushnikov, V. N., Stepokurova, N. A. and Zhukov, S. Ja. (1975). Auditory processing of steady-state vowels. In Fant, G. and Tatham, M. A. A. *Auditory analysis and perception of speech*. London: Academic, Pp. 37–53.

Klatt, D. H. (1980a). Speech perception: A model of acoustic-phonetic analysis and lexical access. In Cole, R. (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J. Lawrence Erlbaum Associates.

Klatt, D. H. (1980b). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971–995.

Lummis, R. C. and Guttman, N. (1972). Exploratory studies of Zwicker's "Negative Afterimage" in hearing. *Journal of the Acoustical Society of America*, **51**, 1930–1944.

McAdams, S. (1980). Spectral fusion and the creation of auditory images. In M. Clynes (Ed.), *Music, mind and brain: The neuropsychology of music*. New York: Plenum Press.

McAdams, S. and Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, 3(4), 26–43, 60.

Miller, G. A. and Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, **22**, 637–638.

Neelen, J. J. M. (1967) Auditory afterimages produced by incomplete line spectra. *IPO Annual Progress Report*, Vol. 2. Eindhoven, Netherlands; Institute for Perception Research. Pp. 37–43.

Rasch, R. A. (1978). Perception of simultaneous notes such as in polyphonic music. *Acustica*, **40**, 21–33.

Summerfield, A. Q., Foster, J., Gray, S. and Haggard, M. (1981). Perceiving vowels from 'flat spectra'. *Journal of the Acoustical Society of America*, **69**, S 116(A).

Schwid, H. A. and Geisler, C. D. (1982). Multiple reservoir model of neurotransmitter release by a cochlear inner hair cell. *Journal of the Acoustical Society of America*, 72, 1435–1440.

Viemeister, N. F. (1980). Adaptation of masking In G. van den Brink and F. A. Bilsen (Eds.), *Psychophysical, physiological and behavioural studies in hearing*. Delft: Delft University Press.

Viemeister, N. F. and Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *Journal of the Acoustical Society of America*, 71, 1502–1507.

Wilson, J. P. (1970). An auditory after-image. In R. Plomp and G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing*. Leiden: Sijthoff, 303–315.

Zwicker, E. B. (1964). An afterimage in hearing. *Journal of the Acoustical Society of America*, 36, 2413–2415.

Manuscript received 18 July 1983